CONCISE COMMUNICATION

# Implementing Automated Surveillance for Tracking *Clostridium difficile* Infection at Multiple Healthcare Facilities

Erik R. Dubberke, MD, MSPH;[1]
Humaa A. Nyazee, MPH;[1] Deborah S. Yokoe, MD;[2]
Jeanmarie Mayer, MD;[3] Kurt B. Stevenson, MD, MPH;[4]
Julie E. Mangino, MD;[4] Yosef M. Khan, MD;[4]
Victoria J. Fraser, MD[1]
for the Centers for Disease Control and
Prevention Epicenters Program

Automated surveillance using electronically available data has been found to be accurate and save time. An automated *Clostridium difficile* infection (CDI) surveillance algorithm was validated at 4 Centers for Disease Control and Prevention Epicenter hospitals. Electronic surveillance was highly sensitive, specific, and showed good to excellent agreement for hospital-onset; community-onset, study facility–associated; indeterminate; and recurrent CDI.

*Infect Control Hosp Epidemiol* 2012;33(3):305-308

It is recommended that all US hospitals track *Clostridium difficile* infection (CDI).[1] At a minimum, it is recommended to conduct surveillance for hospital-onset CDI, but tracking CDI with onset in the community may have important epidemiological and prevention implications.[1,2] However, surveillance for community-onset CDI is much more labor intensive than for hospital-onset CDI. Due to increased demand for patient safety coupled with an emphasis to adopt and implement electronic health records, automated surveillance systems for tracking nosocomial infections needed to be investigated to maximize both limited resources and patient safety.[3,4] The goal of this study was to develop and validate an automated CDI surveillance algorithm using electronically available data at multiple healthcare facilities.

## METHODS

The study population included all adult patients ≥18 years of age admitted to 4 US hospitals participating in the Centers for Disease Control and Prevention (CDC) Epicenters Program from July 1, 2005, to June 30, 2006. These hospitals included Barnes-Jewish Hospital (St. Louis, MO), Brigham and Women's Hospital (Boston, MA), Ohio State University Medical Center (Columbus, OH), and University Hospital (Salt Lake City, UT).

A conceptual automated CDI surveillance algorithm was created by using recommended surveillance definitions (Figure 1).[1] Each center worked with its medical informatics de-partments to apply the algorithm to its local databases. CDI case categorizations by the algorithm were compared to categorizations previously determined by chart review.[5] A second chart review was performed for discordant results. The gold standard comparison was all concordant cases and the categorization determined to be correct by the rereview. The algorithms were modified as needed to improve accuracy. Sensitivities and specificities were calculated for each CDI surveillance definition. Kappa ($\kappa$) statistics were also calculated. Statistical analyses were performed with SPSS for Windows, version 19.0 (SPSS).

## RESULTS

There were 1,767 patients with stool that tested positive for *C. difficile* toxins identified. After the initial comparison of the algorithm's categorization of CDI cases to categorizations determined by chart review, hospital A had 204 discordant cases (27.1%), hospital B had 77 cases (18.7%), hospital C had 55 cases (22.4%), and hospital D had 104 cases (29.1%). Data on discordant cases were submitted back to the appropriate hospitals for rereview.

The overall sensitivities, specificities, and $\kappa$-values of the algorithm by CDI onset compared to the gold standard were as follows: hospital onset: 92%, 99%, and 0.90; community onset, study facility–associated: 91%, 98%, and 0.84; community onset, other healthcare facility associated: 57%, 99%, and 0.65; community onset, community associated: 96%, 94%, and 0.69; indeterminate cases: 80%, 98%, and 0.76; and recurrent cases: 94%, 99%, and 0.94 (Table 1). Similar sensitivity, specificity, and $\kappa$-values were seen at all individual hospitals for community-onset, study-center–associated, and recurrent CDI (Table 1). The algorithm had excellent agreement for hospital-onset CDI at each hospital—except for hospital B. Community-onset and other healthcare facility–associated CDI showed a wide range of sensitivities (16%–96%) and $\kappa$-values (0.25–0.93). Similar trends were seen for community-onset, community-associated, and indeterminate CDI.

Each hospital had to individualize the algorithm to its facility. Hospitals A, B, and C did not have discrete data on where a patient was admitted from (eg, admit from home or a long-term care facility), whereas hospital D did. Therefore, categorization of community-onset cases at these hospitals was dependent on the discharge status (eg, discharge to home or long-term care facility) if the patient had a prior hospitalization in the previous 12 weeks. Hospital A has a code for patients with frequent outpatient visits called "recurring patients," which has a start date of the first visit and end date of December 31. Many recurring patients with CDI were misclassified as hospital-onset CDI. The medical informatics team created a new table within the database that contained
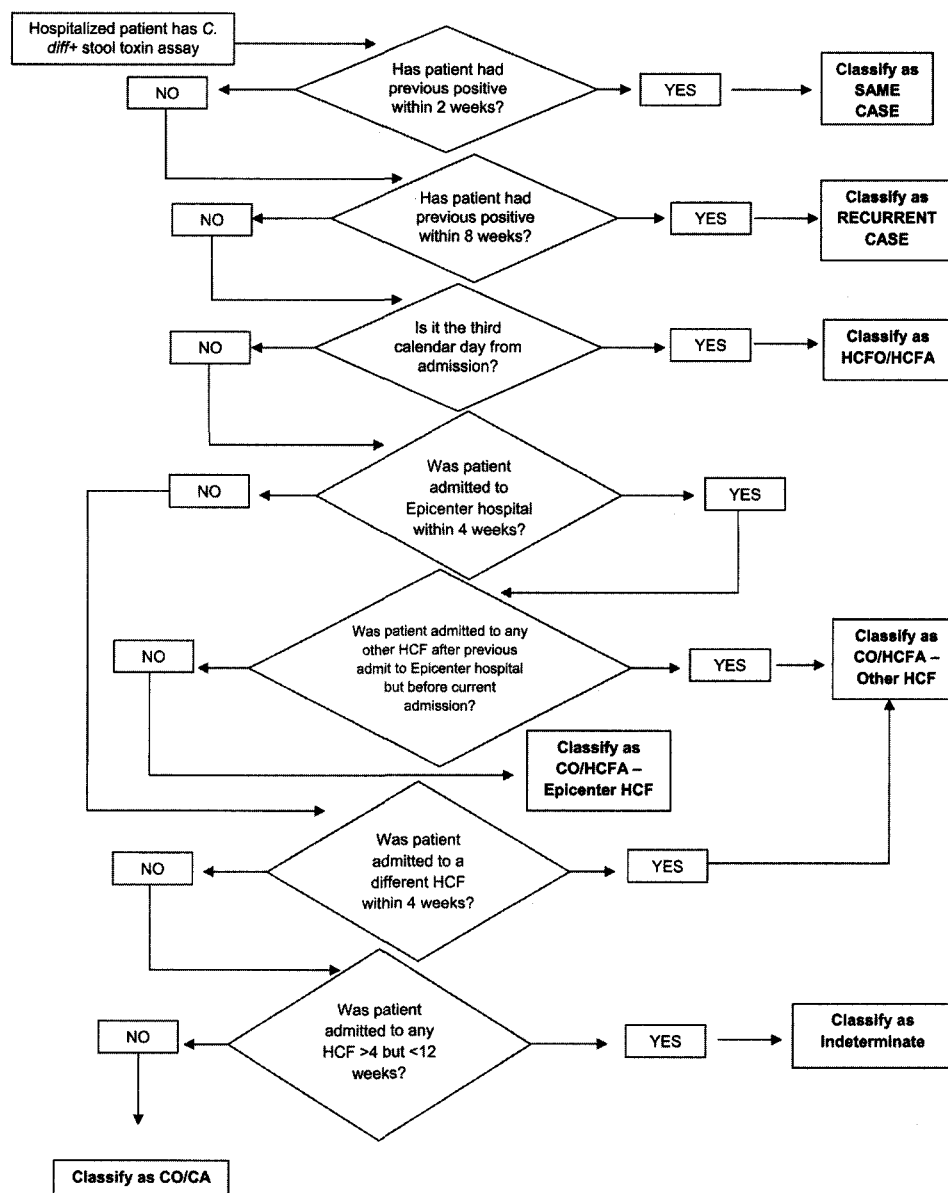
FIGURE 1. Conceptual automated CDI surveillance algorithm.

information regarding the visit type associated with a given encounter to correct this problem. Hospital B made minor modifications to the hospital-onset time cutoff to improve accuracy. Hospital C was not able to modify its algorithm because some data were available only through free text fields. Hospital D initially included patients who were admitted to only 1 particular building, missing those patients who were admitted to the other 3 buildings of their medical center. This was corrected. Three other issues were identified and resolved after the initial review of discordant cases: outpatient encounters were included when determining case categorization rather than only inpatient encounters; only the first positive *C. difficile* toxin result per patient was evaluated, so subse-

quent episodes of CDI were missed; and stool collection date was used to identify patients instead of the admit date.

## DISCUSSION

The goal of this study was to develop and validate an automated CDI surveillance algorithm using existing electronically available data. Previous research indicates that electronic surveillance is more accurate and reliable than manual surveillance.[3,6] Automated surveillance also requires less time because it eliminates the need to do chart review. This study found automated CDI surveillance to be feasible and reliable with overall good to excellent agreement for hospital-onset;

TABLE 1.    Sensitivities, Specificities, and κ-values by CDI Onset and Facility

| Case definition | Hospital[a] | | | | Total |
| --- | --- | --- | --- | --- | --- |
| | A | B | C | D | |
| Healthcare facility onset | 99, 98 (.97) | 75, 99 (.66) | 94, 100 (.93) | 100, 99 (.99) | 92, 99 (.90) |
| Community onset, study center associated | 93, 96 (.83) | 100, 97 (.86) | 84, 99 (.83) | 81, 100 (.88) | 91, 98 (.84) |
| Community onset, other healthcare facility associated | 16, 99 (.25) | 82, 98 (.61) | 53, 98 (.59) | 96, 98 (.93) | 57, 99 (.65) |
| Community associated, community onset | 91, 95 (.71) | 100, 87 (.63) | 100, 92 (.44) | 100, 99 (.91) | 96, 94 (.69) |
| Indeterminate | 83, 98 (.80) | 73, 98 (.63) | 63, 97 (.48) | 84, 99 (.88) | 80, 98 (.76) |
| Recurrent | 99, 99 (.97) | 88, 99 (.85) | 64, 100 (.77) | 97, 100 (.98) | 94, 99 (.94) |

[a] Data are percentages of sensitivity, specificity (κ-value).

community-onset, study facility–associated; indeterminate; and recurrent CDI case categorizations.

Hospitals worked with their individual information technology teams to apply the general automated CDI surveillance algorithm to the data available at their facilities. In this study, data availability and type of data varied from hospital to hospital, thus impacting the accuracy of the automated algorithm. This issue is illustrated by hospital D, the hospital that performed the best at categorizing community-onset CDI because there was a discrete variable that captured where patients were admitted from.

There are potential limitations to the use of an automated CDI surveillance algorithm. Electronic surveillance requires access to an electronic health record (EHR) system. Only ~12% of US hospitals have an EHR system.[7] To develop an automated algorithm, surveillance rules need to be specified into electronic algorithm rules. This can lead to algorithms that vary from site to site based on data availability. As a result, each center can potentially have different rules for the same infection, resulting in different rates, making interhospital comparisons difficult.[3]

Another limitation of using an automated CDI surveillance algorithm is that chart review is not performed. Although the lack of chart review is mitigated by enforcing toxin testing of only diarrheal stool, misclassification is still possible. It is possible that a true community-onset CDI case could be misclassified as a hospital-onset CDI case if stool were collected after the hospital-onset cutoff date. In addition, patients with a positive assay for *C. difficile* may not have clinically significant diarrhea and therefore do not truly have CDI. This may be especially problematic at hospitals that use nucleic acid amplification tests.[8]

This study found automated electronic CDI surveillance to be highly sensitive and specific for identifying cases of hospital-onset; community-onset, study center–associated; and recurrent CDI. Automated CDI surveillance will allow infection preventionists to devote more time to infection prevention efforts. In addition, automated CDI surveillance may facilitate a healthcare facility's ability to track community-

onset CDI. Community-onset CDI likely contributes to hospital-onset CDI because patients admitted to a healthcare facility with CDI are a source of *C. difficile* transmission to other patients. Understanding the burden of community-onset CDI may allow for targeting of CDI prevention efforts.[2] Implementing an automated algorithm using electronically available data is feasible and reliable.

REFERENCES

1. McDonald LC, Coignard B, Dubberke E, et al. Recommendations for Surveillance of *Clostridium difficile*–associated disease. *Infect Control Hosp Epidemiol* 2007;28:140–145.

2. Dubberke ER, Butler AM, Hota B, et al. Multicenter study of the impact of community-onset *Clostridium difficile* infection on surveillance for *C. difficile* infection. *Infect Control Hosp Epidemiol* 2009;30:518–525.

3. Hota B, Lin M, Doherty JA, et al. Formulation of a model for automating infection surveillance: algorithmic detection of central-line associated bloodstream infection. *J Am Med Inform Assoc* 2010;17:42–48.

4. Centers for Disease Control and Prevention. 2008. Multidrug-resistant organism (MDRO) and *Clostridium difficile*–associated disease (CDAD) module [PowerPoint Slides]. Retrieved from http://www.cdc.gov/nhsn/wc_MDRO_CDAD_ISlabID.html.

5. Dubberke ER, Butler AM, Yokoe DS, et al. Multicenter study of *Clostridium difficile* infection rates from 2000 to 2006. *Infect Control Hosp Epidemiol* 2010;31:1030–1037.

6. Klompas M, Kleinman K, Platt R. Development of an algorithm for surveillance of ventilator-associated pneumonia with electronic data and comparison of algorithm results with clinician diagnoses. *Infect Control Hosp Epidemiol* 2008;29:31–37.

7. Jha AK, DesRoches CM, Kralovec PD, Joshi MS. A progress report on electronic health records in U.S. hospitals. *Health Aff* 2010; 29:1951–1957.

8. Dubberke ER, Han Z, Bobo l, et al. Impact of clinical symptoms on interpretation of diagnostic assay for *Clostridium difficile* infections. *J Clin Microbiol* 49:2887–2893.